DNA Bubbles

Kim Ø. Rasmussen (T-11), N.K. Voulgarakis (University of Crete, Greece), S. Ares (Universidad Carlos III de Madrid, Spain), and Alan R. Bishop (T-DO); ***kor@lanl.gov***

Accessing the genetic code stored in DNA is central to fundamental biological processes such as replication and transcription, and this requires that the extraordinarily stable double-helical structure of the molecule must locally open to physically expose the bases. Although, in a cell proteins may play a role in creating and sustaining local strand separations, recent evidence corroborates that thermal effects play an extremely important role.

The actual melting of double-stranded DNA occurs through an entropy-driven phase transition. The entropy gained in transitioning from the unusually rigid double-stranded DNA to the much more flexible single-stranded DNA can, already at moderate temperatures, balance the energy cost of breaking a base-pair. Since the double-stranded helix is held together by hydrogen bonds between complementary base-pairs—two bonds for the AT pair and three bonds for the stronger GC pair—the sequence heterogeneity interplays with the entropy effects to create an extended premelting temperature window, (including the biologically relevant regime) where large thermal bubbles are readily formed.

On the experimental side the properties of denaturation bubbles can now be studied *in vitro* and provide important insight to the biological processes. Recent, experimental studies [1] have attempted to interrogate the nature and statistical significance of such bubble states. Intriguingly, these experiments combine traditional UV absorption experiments with a novel bubble-quenching technique that traps ensembles of bubbles to capture statistical properties of the bubbles. The experimental studies concentrated on short DNA sequence in which a soft sequence (AT rich) of varying length was embedded in a more tightly bound (GC rich) sequence of fixed length. With this design the bubbles are expected to form primarily in the soft section of the molecule and this allowed a detailed study of sequence length and temperature dependence.

In this work we compare these powerful recent experimental results with Monte Carlo simulations of the model proposed by Peyrard, Bishop, and Dauxois (PBD). The PBD model has already been successfully compared with denaturation experiments on short homogeneous sequences. More

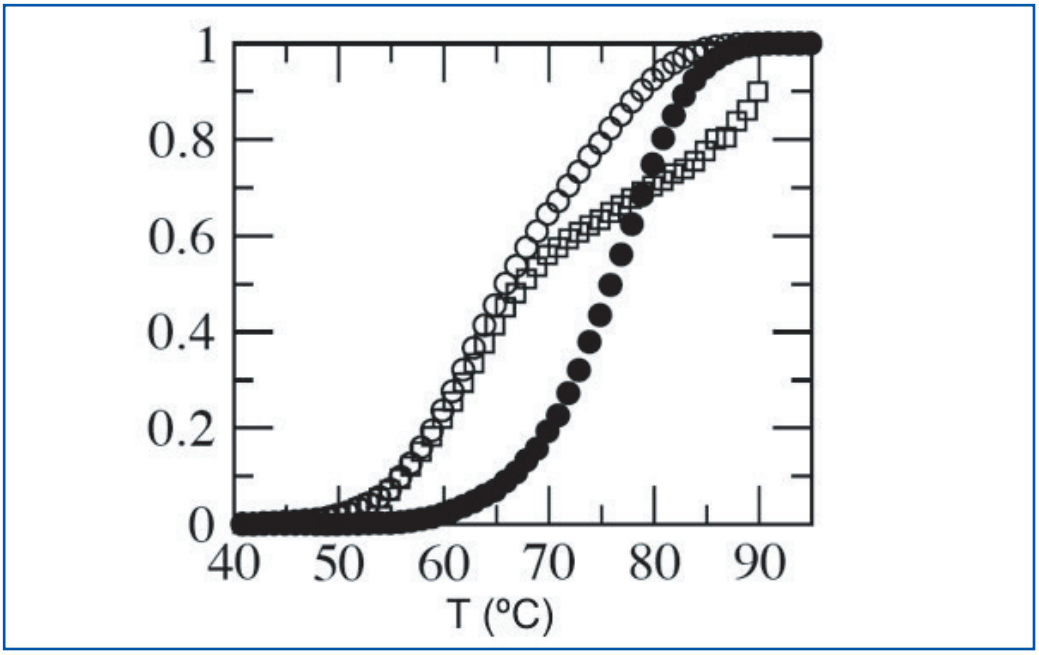

Figure 1—
Results of Monte Carlo simulation of the PBD model encoding a 60-base-pair sequence with a soft core of 36 AT base pairs. Open circles indicate the fraction of open bases, while filled circles show the fraction of completely denaturated (melted) molecules versus temperature. By combining these results one can determine the fraction (indicated by squares) of base pairs participating in a bubble (the size of a bubble). These results accurately reproduce the experimental results on the same molecule (see Ref. [1]).

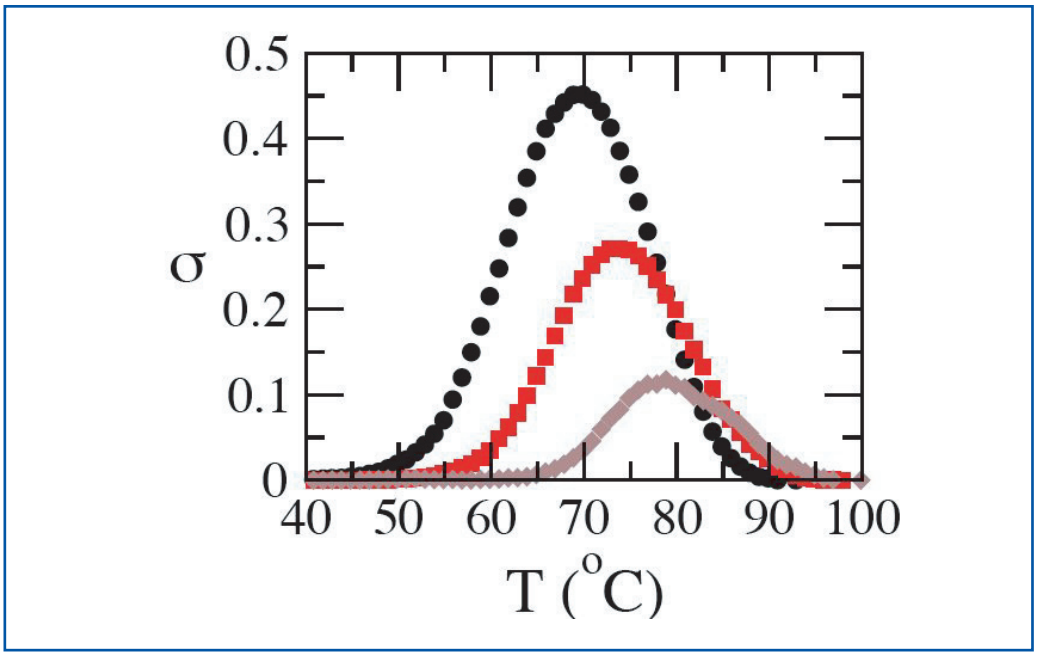

Figure 2— Bubble size (σ as a fraction of full molecule) versus temperature for various DNA sequences lengths L=60, (filled triangles), 42 (red squares) and 33 (brown diamonds).

exceptional is the recent [2] accurate prediction of the location at which large bubbles form in several viral sequences and how those locations relate to transcription initiation sites. The difference between our comparison and previous ones is that we use the same (deceptively) simple model, with no further refinements that introduce new parameters that need to be fitted. Indeed parameters of the model are not changed to fit the experiments: we use the *same* values for those parameters that were fixed for a different purpose. Figure 1 shows the result of such a simulation on a 60-base-pair DNA molecule, which consists of 36 AT base pairs framed on each side by 12 GT base pairs. The open circles show the fraction of open base pairs versus temperature. At the same temperature the filled circles indicate the faction of completely melted molecules. From these results one can determine the fraction of base pairs that participate in a bubble as shown by squares. It is clear that there is a significant temperature window in which part of this DNA molecule can be open.

To further investigate this we show in Fig. 2 the fraction σ of bases participating in bubbles versus temperature for three different sequences. The black circles are for the sequence in Fig. 1, i.e., 36 TA base framed between 12 GC base pairs. The red and the brown squares are for 18 and 9 base pairs again framed between 12 GC base pairs respectively. It is noticeable that the maximum bubble never occupies the entire "soft" fraction (0.6, 0.43, and 0.27 respectively) of the molecule but always a somewhat smaller fraction (0.45, 0.25, and 0.1 respectively). Also noteworthy is that the maximum bubble occurs at an increasingly higher temperature as the "soft" faction of the molecule becomes smaller. It is important to stress that these results accurately reproduce the experimental observations of Ref. [1] on the same sequences.

[1] Y. Zeng, A. Montrichok and G. Zocchi, *J. Mol. Biol.* **339**, 67 (2004).
[2] C.H. Choi, et al., *Nucleic Acids Res.* **32**, 1584 (2004).